Networking and Information Technology Research and Development
NITRD - A Testing Perspective

New applications over computer networks appear every day ranging from simple message systems to VOIP and IPTV. Also, a multitude of new protocols for ad hoc wireless network have been proposed. Despite all the improvements on network communication systems, little has been done to advance the way these new applications/protocols are tested. Most existing work on network testing is restricted to protocol conformance and many criteria have been proposed to assess the quality of a test suite. However, those criteria focus on the node level and not on the network level. For example, state coverage is used to assess how many states of a protocol have been exercised by a determined test suite. The same applies to transition coverage. While these criteria are important, they do not provide an overall indication about the quality of a test suite at the network level. Some criteria have been proposed to account for all possible interactions between a communication protocol for end-to-end nodes. While improving over state and transition coverage, they do not provide assessment criteria at the network level. If testing an isolated node or a couple of nodes was sufficient, then there would be no need for large test beds such as DETER, PlanetLab and simulation systems such as ns2, where experiments with thousands of nodes can be conducted.

One question to be answered when testing any system is “How adequate is the test suite?”, that is “How can researchers qualify their network experiments?” If the adequacy of a test suite cannot be properly measured, then it is very unlikely that someone will have high confidence in the quality of the product to be released. The same applies for network testing where a network experiment or set of experiments is analogous to a test suite. Basically, three measurements have been used to determine the adequacy of a test suite for network testing: scalability; execution time; and confidence interval. A myriad of research papers and industrial experiments use these factors as the main aspects to evaluate their test suites. These factors form an important subpart but do not constitute the entire picture of test suite adequacy. For example, suppose an experiment is conducted using a thousand nodes executing over a period of five days (120 hours) using some appropriate random traffic generator. Would such an experiment be “adequate” for exercising the network? In general, there can be no definitive answer. Further, the same events could be generated over and over again and many possible scenarios could be left untouched by the experiment. The experiment may be repeated many times and a 99% confidence interval is computed. Even in this case, there is no guarantee that the experiments are not biased. This is a clear indication of the neccessity of better testing tools and techniques for networking experiments. The identification of multi-layer or cross-layer assessment criteria along with automatic instrumentation of source code with respect to the identified criterion would bring the current practise in networking testing to a

much higher level. Although one cannot validate other's experimental results the availability of a proper assessment criteria gives an indication of how much one can trust the produced results.

Also, not many experiments are conducted using test beds such as DETER or PlanetLab. One of the main reasons is the effort require to deploy and execute an experiment on these test beds. There is a clear need for tools that can help on the (re)deployment of network experiments. The availability of such tools would encourage the testing of proposed networking solutions using more real life environments and would consequently improve their quality. The tool should also facilitate the reconfiguration of the environment to make possible the testing of more diverse scenarios.

Another problem to be addressed is to avoid unnecessary resource usage. Many experiments are programmed to run for many days or even weeks but they may reach a saturation point (no progress in terms of testing new scenarios; rate of change of coverage is zero) earlier in this period. In this case, all the rest of the time and resources are being wasted as no new scenarios are tested. Therefore techniques to identify the "saturation point" as well as to dynamically tune the experiments to regain the testing of new scenarios would result in substantial savings in terms of time (no need to abort and re-start the experiment) and resource usage. This goal can be achieved by: (i) dynamically modeling the behavior of the networking experiment; (ii) automatically identifying the inputs and variables affecting the resource under observation; and (iii) tuning these inputs to regain increase in coverage.

In summary, the following itens are needed to improve the current status of network testing:

- a tool to facilitate deployment of experiments on easily reconfigurable test beds.
- a cross-layer or multi-layer network assessment criteria and correspondent coverage analysis tools.
- a technique to identify the saturation points during execution of experiments and dynamically tune the experiments.

A multi-agency collaboration is keen to realize the proper development of the testing tools/techniques listed above. First, deployment tools may need to follow a specific standard or language. The individual development of them by each agency would lead to a myriad of distinct languages and would defy the goal of easy deployment in any given scenario. Second, assessment criteria defined according to specific needs of an agency would be weaker than a more comprehensive set of criteria defined in a global manner. Third, some agencies may have a focus on some specific networking needs and the combined effort would lead to the development of tools with a broader scope and consequently more effective testing tools. Finally, testing tools are general purpose and the multi-agency development would incur substantial savings.